IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Muhammad Hashim
October 1, 2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- In this report data from SpaceX was used to predict whether a future launch would be successful based on its cargo mass, launch site and other factors.
- Data was collected using SpaceX API and scrapped from public records

Introduction

- The purpose of the analysis is to predict if the Falcon 9 first stage will land successfully. The main selling point of SpaceX is that they can reuse the first stage saving millions per launch. Therefore if we can determine if the first stage will land, we can determine the cost of a launch.

Very Expensive

Also very expensive... but a few million are saved

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and extracted from public records of Falcon9 Launches
- Perform data wrangling
  - Data was classified into successes and failures according to outcome
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

Collection was done in 2 steps, first using SpaceX's API and then web scrapping from Wikipedia to add more data. When data was complete it was pre-processed to allow for further study.

Sources:
Applied-Data-Science-Capstone/01jupyter-labs-spacex-data-collection-api.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)

Applied-Data-Science-Capstone/02jupyter-labs-webscraping.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)

Applied-Data-Science-Capstone/03labs-jupyter-spacex-Data wrangling.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)

API dataframe

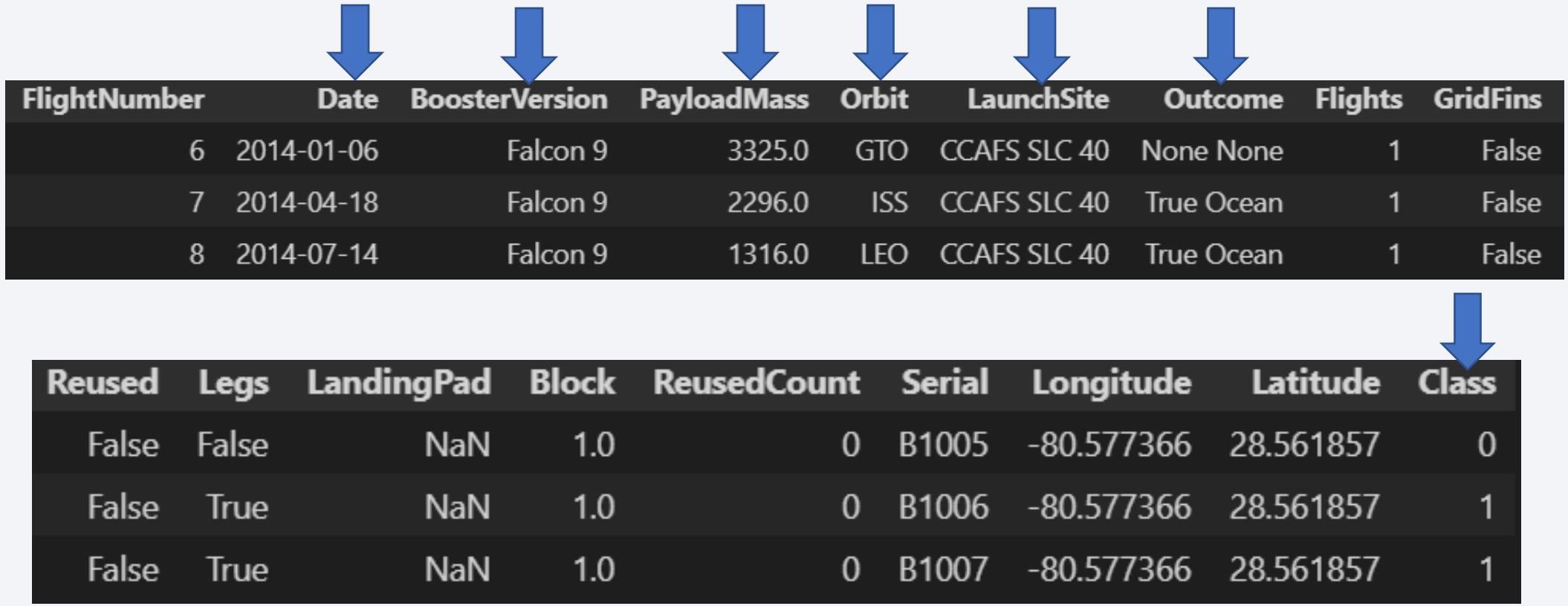
FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights
1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1
2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1
3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1
4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1
5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1

GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
False	False	False	None	1.0	0	B0003	-80.577366	28.561857
False	False	False	None	1.0	0	B0005	-80.577366	28.561857
False	False	False	None	1.0	0	B0007	-80.577366	28.561857
False	False	False	None	1.0	0	B1003	-120.610829	34.632093
False	False	False	None	1.0	0	B1004	-80.577366	28.561857

Scrapped data

Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
7	CCAFS	SES-8	3,170 kg	GTO	SES	Success	F9 v1.1	No attempt	3 December 2013	22:41
8	CCAFS	Thaicom 6	3,325 kg	GTO	Thaicom	Success	F9 v1.1	No attempt	6 January 2014	22:06
9	Cape Canaveral	SpaceX CRS-3	2,296 kg	LEO	NASA	Success\n	F9 v1.1	Controlled	18 April 2014	19:25
10	Cape Canaveral	Orbcomm-OG2	1,316 kg	LEO	Orbcomm	Success	F9 v1.1	Controlled	14 July 2014	15:15

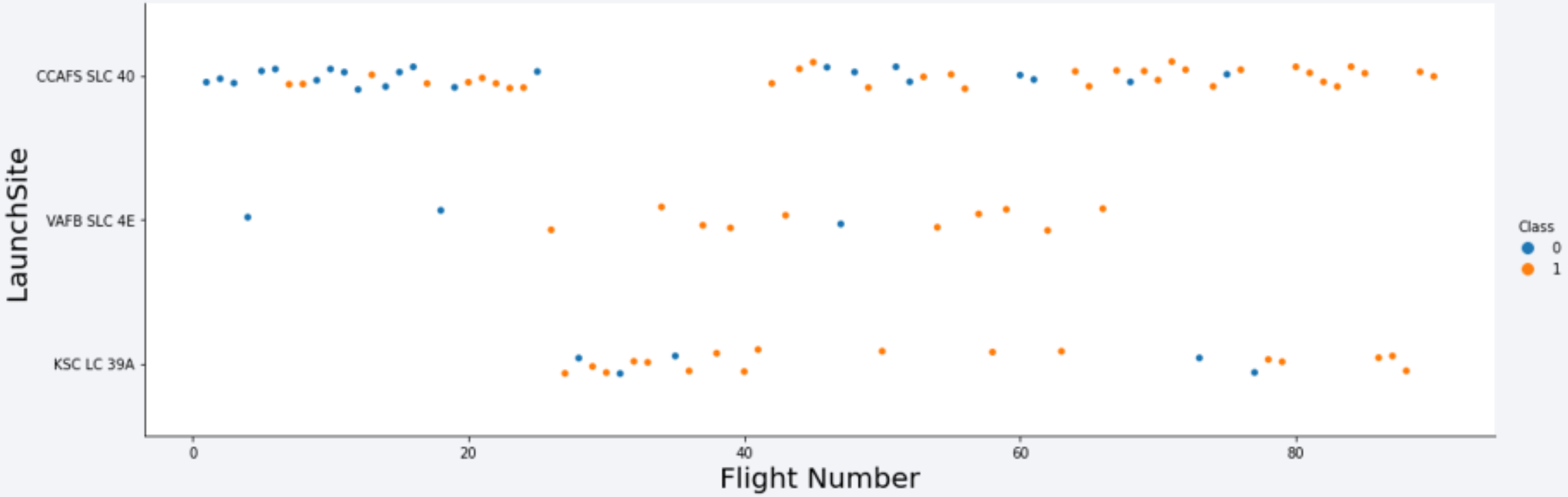
Data Wrangling



FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins
6	2014-01-06	Falcon 9	3325.0	GTO	CCAFS SLC 40	None None	1	False
7	2014-04-18	Falcon 9	2296.0	ISS	CCAFS SLC 40	True Ocean	1	False
8	2014-07-14	Falcon 9	1316.0	LEO	CCAFS SLC 40	True Ocean	1	False

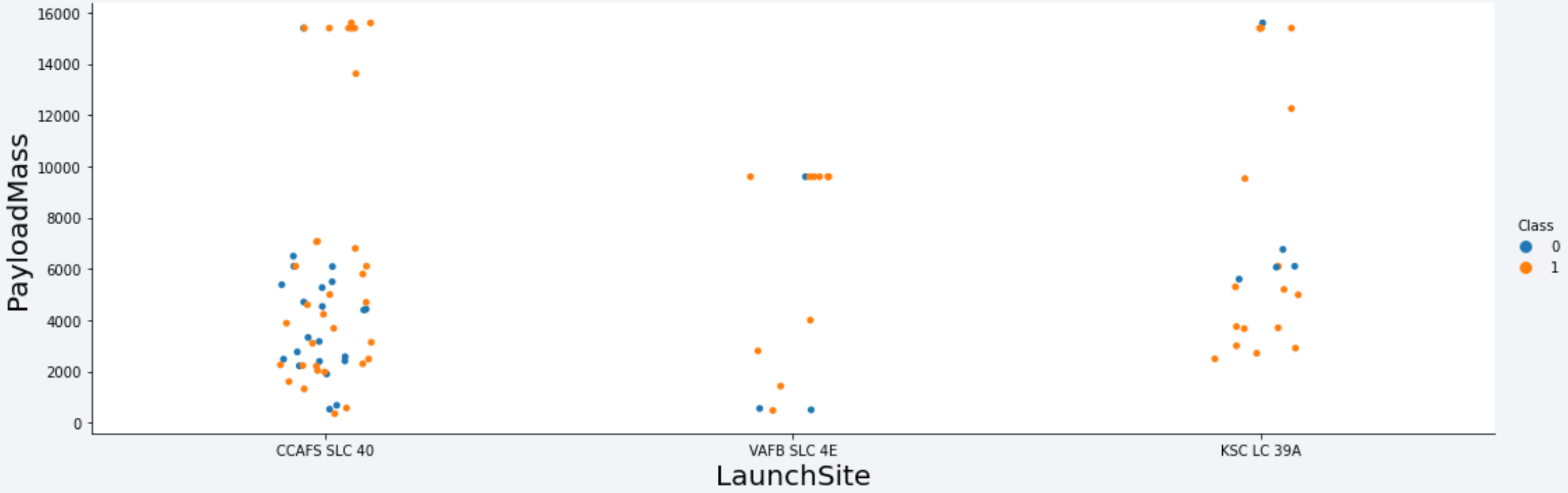
Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
False	False	NaN	1.0	0	B1005	-80.577366	28.561857	0
False	True	NaN	1.0	0	B1006	-80.577366	28.561857	1
False	True	NaN	1.0	0	B1007	-80.577366	28.561857	1

EDA with Data Visualization



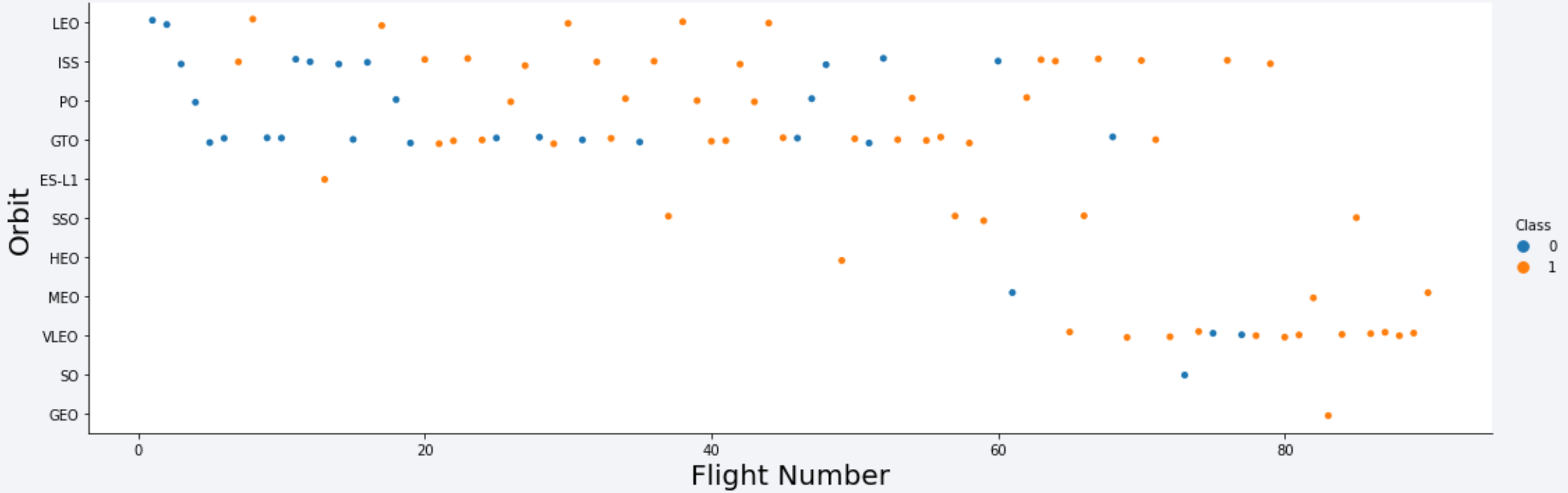
[Applied-Data-Science-Capstone/05EDA with Visualization lab.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)]

EDA with Data Visualization



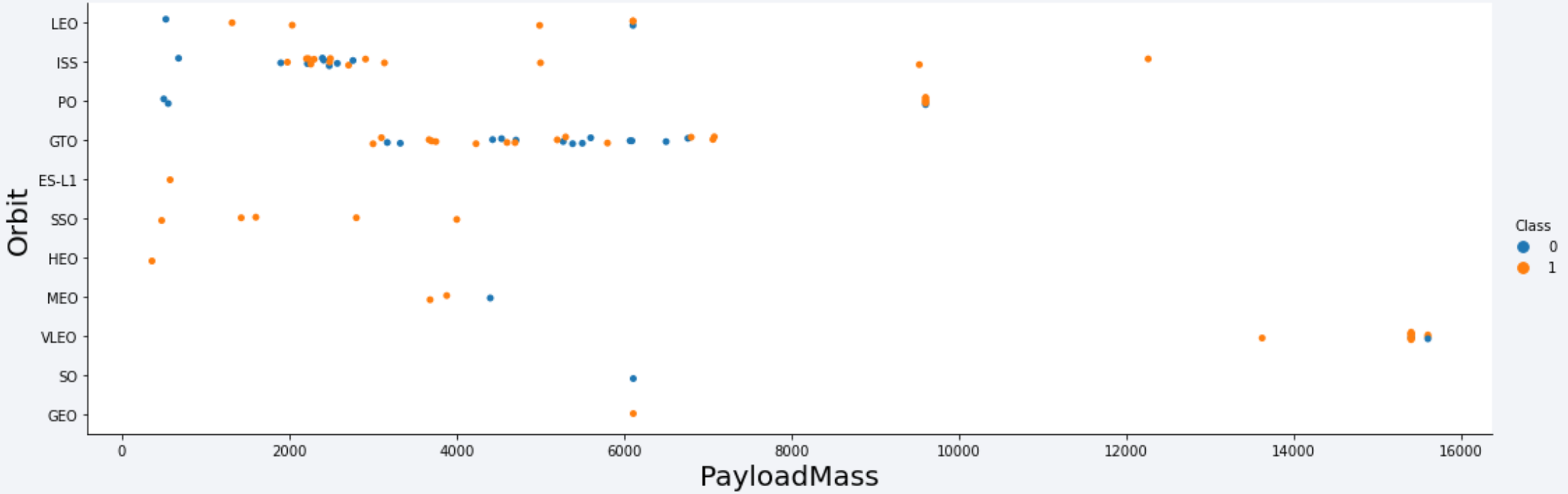
[Applied-Data-Science-Capstone/05EDA with Visualization lab.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)]

EDA with Data Visualization



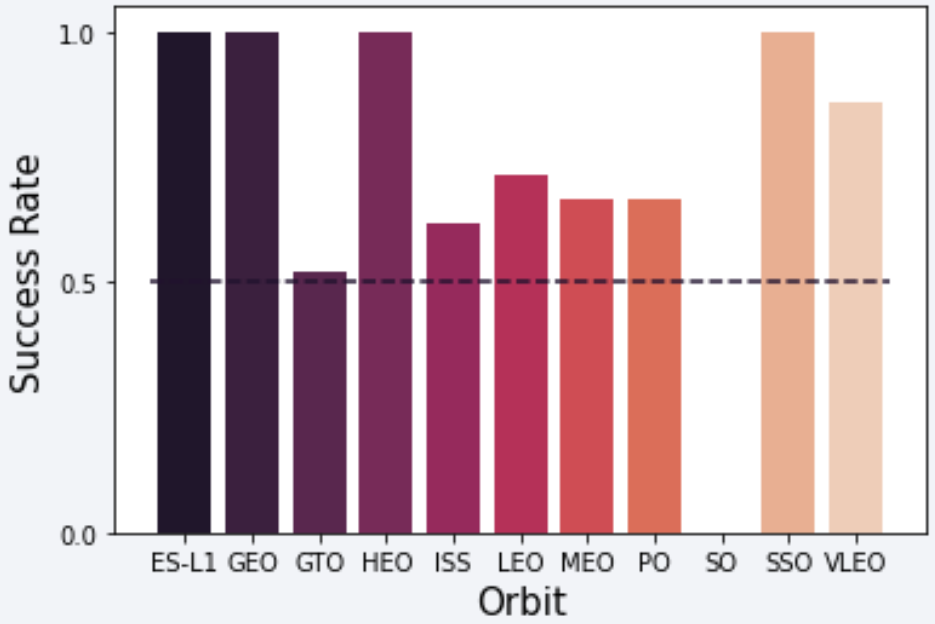
[Applied-Data-Science-Capstone/05EDA with Visualization lab.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)]

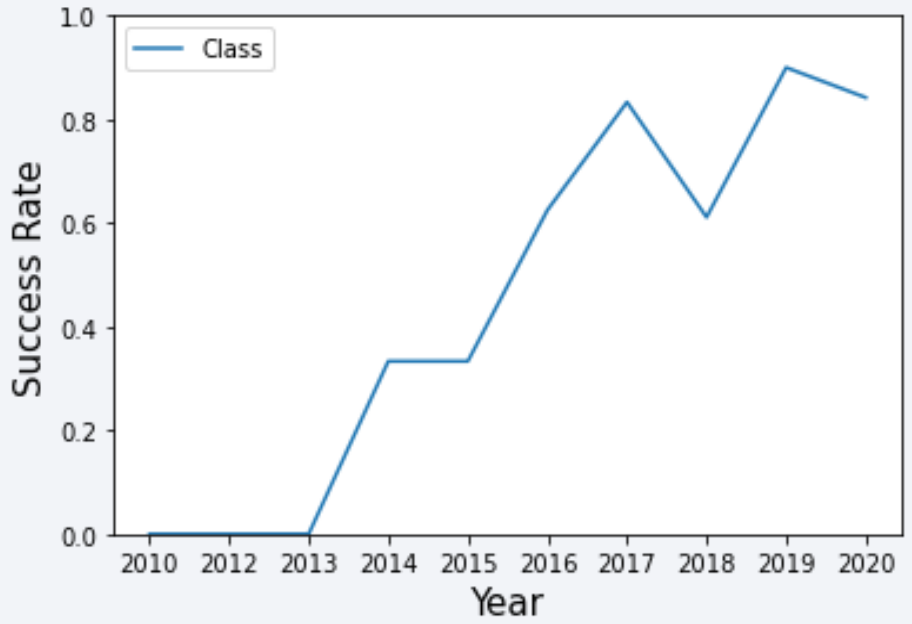
EDA with Data Visualization



[Applied-Data-Science-Capstone/05EDA with Visualization lab.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)]

EDA with Data Visualization





[Applied-Data-Science-Capstone/05EDA with Visualization lab.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)]

Section 5

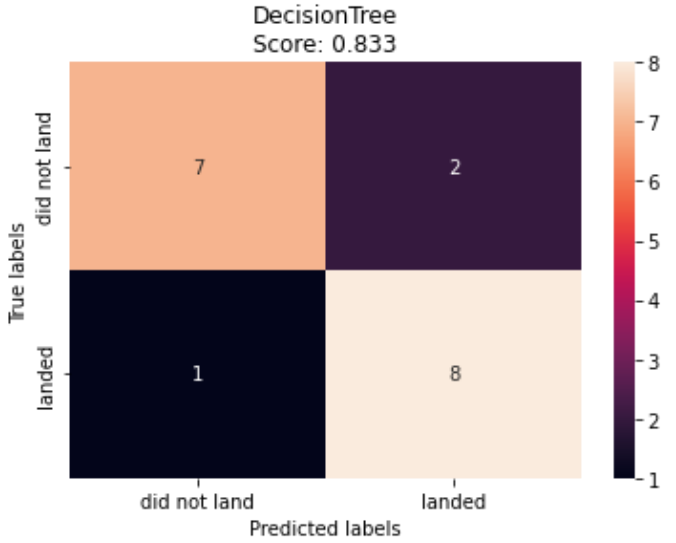
Predictive Analysis (Classification)

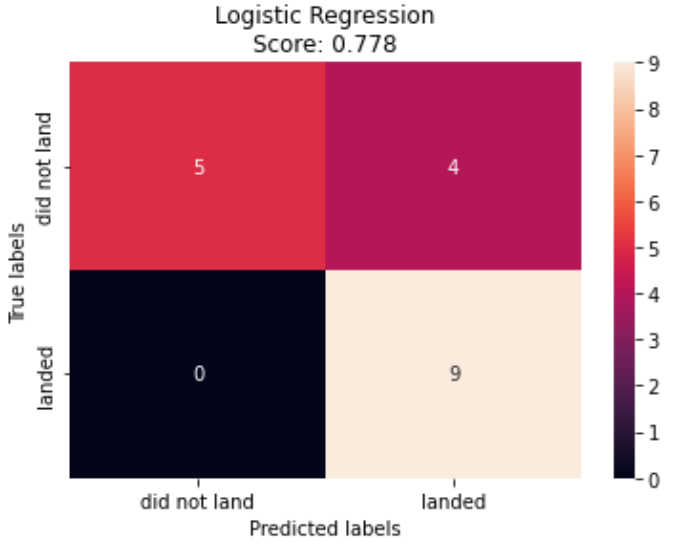
Predictive Analysis (Classification)

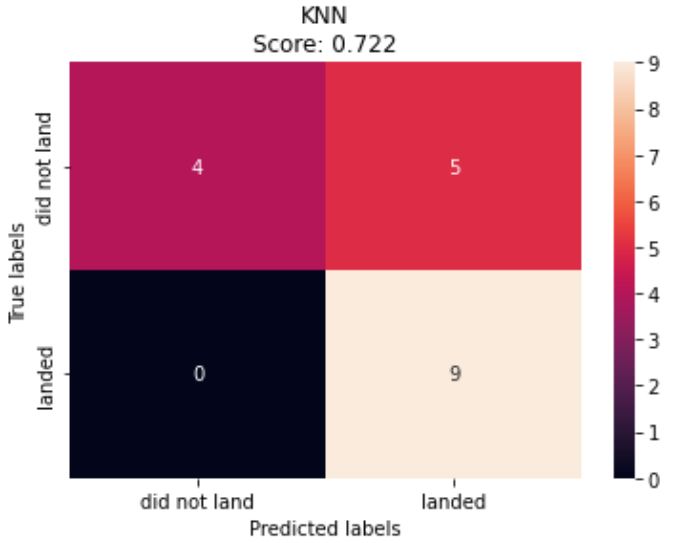
Several classification models to predict launches:

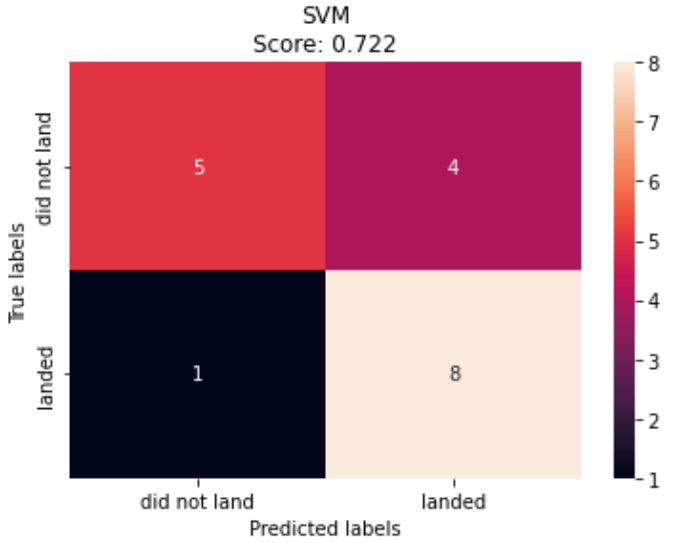
- Logistic Regression
- SVM
- Decision Tree
- KNN

Classification Accuracy









[Applied-Data-Science-Capstone/08SpaceX_Machine Learning Prediction_Part_5.ipynb at master · JP235/Applied-Data-Science-Capstone (github.com)]

Thank you!